

Opening the black box of data-based school monitoring: Data infrastructures, flows and practices in state education agencies

Big Data & Society
January–June 2019: 1–12
© The Author(s) 2019
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/2053951719853311
journals.sagepub.com/home/bds



Sigrid Hartong and Annina Förschler

Abstract

Contributing to a rising number of *Critical Data Studies* which seek to understand and critically reflect on the increasing datafication and digitalisation of governance, this paper focuses on the field of school monitoring, in particular on digital data infrastructures, flows and practices in state education agencies. Our goal is to examine selected features of the enactment of datafication and, hence, to open up what has widely remained a black box for most education researchers. Our findings are based on interviews conducted in three state education agencies in two different national contexts (the US and Germany), thus addressing the question of how the datafication and digitalisation of school governance has not only manifested within but also across educational contexts and systems. As our findings illustrate, the implementation of data-based school monitoring and leadership in state education agencies appears as a complex entanglement of very different logics, practices and problems, producing both new capabilities and powers. Nonetheless, by identifying different types of 'doing data discrepancies' reported by our interviewees, we suggest an analytical heuristic to better understand at least some features of the multifaceted enactment of data-based, increasingly digitalised governance, within and beyond the field of education.

Keywords

Data-based governance, digitalisation, education, state education agencies, Germany, US

Introduction

This paper seeks to contribute to the fast-growing body of *Critical Data Studies* by providing empirical insights into the pursuit of data, measurement and commensuration in the field of public education. As in many other governmental spheres (for a recent overview see Smith, 2018: 3), the growing development of digital data infrastructures in education raises numerous questions '[...] about the nature of data, how they are being produced, organized, analyzed and employed, and how best to make sense of them and the work they do. Critical data studies endeavours to answer such questions' (Kitchin and Lauriault, 2014: 1; see also Iliadis and Russo, 2016), while explicitly challenging the idea of data as being neutral or simply technical.

In fact, there is a visibly growing body of work that describes the expanding datafication and digitalisation

of education policy and practice, enhanced by the promotion of the so-called evidence-based governance (e.g. Bellmann, 2015; Grek and Ozga, 2010), including research that has explicitly focused on the production and processing of international student assessments (Bloem, 2016; Gorur, 2014; Lewis, 2017; Villani, 2018). Nonetheless, the increasingly digital and automated formation, recoding, storage, manipulation and distribution of data, all of which have become integral features of education governance (Hartong, 2016,

Faculty of Humanities and Social Sciences, Department of Education, Helmut Schmidt University, Hamburg, Germany

Corresponding author:

Sigrid Hartong, Faculty of Humanities and Social Sciences, Department of Education, Helmut Schmidt University, Postfach 700822, Hamburg 22008, Germany.

Email: hartongs@hsu-hh.de



2018a; Landri, 2018; Sellar, 2015; Selwyn, 2014: 1; Williamson, 2017), have not yet been extensively examined (see West, 2017 for an important exception), representing a ‘black box’ for most education researchers and practitioners. In other words, as described by Selwyn (2014: 13–14), there remains a pressing need to better understand ‘[...] how various forms of digital data are [specifically] set to work within educational contexts, including what data is used, what the uses and consequences are, and how data has become embedded within different organisational cultures’.

With this paper, we seek to respond to this need by examining selected features of expanding data infrastructures, flows and practices of school monitoring in three state¹ education agencies in two different national contexts, the *Massachusetts Department of Elementary and Secondary Education* (DESE) in the US and the *Hamburg Department for Schools and Vocational Education* (BSB) as well as the *Institute for Educational Monitoring and Quality Improvement* (IfBQ), an institution attached to the BSB, in Germany. In both (federally organised) countries, the past two decades have witnessed either a tremendous turn towards (Germany) or a significant expansion (US) of datafication in education, promoted by a strong national coalition for evidence-based policy, which resulted in an extensive implementation and transformation of data infrastructures and flows (Anagnostopoulos et al., 2013; Hartong, 2016, 2018a). Simultaneously, state education agencies in both countries have been urged to produce growing amounts of data and to use that data for more effective and efficient school leadership and monitoring, particularly but not limited to holding schools accountable for digital data production (González-Sancho and Vincent-Lancrin, 2016; Piattoeva, 2016).² As a result, state education agencies have increasingly focused on and restructured themselves around data production, analysis, management and reporting, thus illustrating what Smith (2018: 3) recently described as the ‘dataism’ paradigm reshaping the everyday business of multiple actors and agencies.³

While the main goal of our study is to unpack for school monitoring what Kitchin and Lauriault (2014) describe as ‘data assemblages at work’, we also seek to contribute to a growing number of studies that focus on how the datafication and digitalisation of educational governance has manifested across educational contexts and systems. Notwithstanding a clear globalness in terms of ongoing transformations and thus broad commonalities between datafication policies in various countries (e.g. Lingard et al., 2015; Williamson et al., 2018), such examinations have also identified the significant influence of local contexts – including cultural, social or institutional settings – resulting in a significantly different ‘re/territorialisation’ of data

infrastructures, flows and practices (Hartong, 2018a). Two of many examples are Schildkamp and Teddlie’s (2008) analysis of School Performance Feedback Systems in the US and the Netherlands, and a comparative study on educational data production, availability and use in China, Russia and Brazil by Centeno et al. (2018). The study presented here complements such analyses of digital technologies sitting ‘alongside pre-existing cultures and structures of educational settings’ (Selwyn, 2013: 209), while simultaneously filling a gap by focusing on a key, yet widely under-researched actor in the digitalisation of education governance so far, namely state education agencies and their role as ‘data hubs’ between global, national and local data infrastructures and flows.

The following section is devoted to further, yet brief, conceptual and methodological explanations before we explore the results of the study, principally drawing on 16 interviews with 20 state agency experts conducted in Hamburg and Massachusetts between December 2017 and April 2018.⁴ Our particular emphasis hereby lies in documenting how the implementation of data-based school monitoring and leadership appears less as a purely technical procedure, but instead as a complex entanglement of very different (technical and social) logics, practices and problems. Specifically, we identify different types of ‘doing data discrepancies’, which, as we discuss in our conclusion, illustrate and conceptualise typical challenges associated with the pursuit of data, measurement and commensuration across many other domains of governmental or state activity, thus also offering important implications for the wider field of critical data studies.

Conceptual and methodological framing

The goal of our study was to better understand how datafication, in particular the ‘doing’ of school monitoring and leadership, has become enacted in two state education agencies across two country contexts. Since we have already discussed this conceptual framing extensively in former contributions (Hartong, 2018a, 2019), this section is limited to a brief summary of the main concepts employed:

A central theme of our analysis is that of *data infrastructures* (or *data assemblages*, as used by Kitchin, 2014) to describe and explore the networks of objects and subjects assembled around the socio-technical de- and recontextualisation of data in education (Anagnostopoulos et al., 2013: 8; Hartong, 2018a: 135–136; Sellar, 2015; Williamson, 2017), here in state education agencies. In other words, attention is drawn to ‘[...] the technological, political, social and economic apparatuses and elements that constitute [...] and frame [...] the generation, circulation and deployment

of data' (Iliadis and Russo, 2016: 2, citing Kitchin and Lauriault, 2014), to trace how data enter or are selected into a particular form, related to each other and, affected by these forms and relations, become information or governing knowledge (Sellar, 2017; Thompson and Sellar, 2018: 4–5). Thus, following Williamson (2017: 38–39) as well as Kitchin (2014), we understand data infrastructures not as static, but as practically enacted and constantly transformed by contingent, relational and contextual discursive and material practices. In the case of state school monitoring, such practices particularly refer to the collection, processing, modelling, management, visualisation or reporting of data, but also to the ways in which data is either problematised or taken for granted.

As our empirical observations will show, a key mechanism within the doing of state school monitoring is the fabrication of commensuration, which is the transformation of different qualities into comparable, usually quantified metrics (Espeland and Stevens, 1998). At the same time, however, commensuration requires enormous organisation, decision-making and weighting which, as many of our interviewees reported, can pose significant challenges (which are mostly externally invisible). The problem of fabricating commensuration equally concerns software/coding activities and the embedding of these kind of activities into wider institutional practices (e.g. school support, accountability or reporting), which, to a large extent, means linking numbers to norms, values, and politics, and vice versa – for example by deciding which targets schools are expected to meet and, consequently, when to intervene as a state. As Diesner (2015) has argued, small decisions can thus produce a big (governmental) impact. Consequently, aside from an earnest attempt to understand the enactment of data infrastructures in state education agencies, we aim to unpack (and typify) at least some of these underlying, often ambivalent and difficult decision-making processes, including their political implications (e.g. profiling, social sorting, control creep, etc., see also Kitchin and Lauriault, 2014).

With our exploration of state education agencies in Massachusetts (US) and Hamburg (Germany), we selected contrasting and simultaneously similar cases, resulting from a complex entanglement of different contextual dimensions (Sobe and Kowalczyk, 2018). On the one hand, the US and Germany contain federal, multi-level architectures, where state education authorities to a large extent decide on the implementation, transformation and use of education monitoring systems. On the other hand, both countries stand in stark contrast in terms of using and relying on (quantified) data for educational governance. While in the former we find a strong traditional belief in the value of testing, rankings and the expertise of private test

providers (Sacks, 1999), the latter has for a long time placed its faith more strongly in teachers, exerting '[...] weak control and evaluation of the processes and almost no external control of the outcomes of schooling' (Hopmann, 2003: 472). Even though Germany underwent a tremendous turn towards data-based school governance at the beginning of the 21st century (thus still qualifying as a global 'latecomer'), this scepticism towards standardised testing and public rankings is still largely visible. In contrast, at least for the last 40 years, educational governance in the US has been characterised by the ever-growing importance of data-based accountability (Schildkamp and Teddlie, 2008: 262), further intensified by the so-called *No Child Left Behind* resolution in 2001 (Anagnostopoulos et al., 2013; West, 2017). While an examination of doing monitoring in state education agencies needs to take into account these wider cultural and political contexts that frame 'national imaginaries' (Sobe and Kowalczyk, 2018) of datafication, the same is true for the different state contexts within both federally shaped countries. In other words, the range of differences between state education agencies' datafication processes *within* the US and Germany may appear even larger than country differences. Responding to this additional complexity, this paper focuses on two states that, in relation to their state peers, present themselves as very advanced and experienced in terms of data-based school monitoring and leadership: Massachusetts and Hamburg. Furthermore, both represent relatively small states in which data collection and centralisation appear less problematic than in more territorially extensive states.

Methodologically, the presented findings draw firstly on material collected through extensive online research and document analysis, including organisation charts, policy papers, documentation on the development and usage of data instruments, as well as online data dashboards. Building on this initial research, we further conducted 16 semi-structured interviews with 20 state agency experts, ranging between 60 and 90 minutes each. We focused on the most relevant institutions conducting state-level data work related to school monitoring and leadership – the DESE in Massachusetts and, for Hamburg, the BSB as well as the IfBQ. We talked to as many 'data experts' as access allowed, operating across the fields of data collection, validation, modelling, storage, processing and distribution.

Having concluded transcription of the interviews, we completed multiple reviews of the collected material, using topical coding as well as conceptual framing outlining (Rivera, 2018: 8). We first reviewed sources for the two cases separately, annotating the text using codes referring to (a) the data infrastructure and flows in educational monitoring and (b) descriptions of specific data

practices. We then combined the annotated text sections from both cases in a new document, sorting, comparing and typifying data infrastructures and practices described across the two cases/three agencies. Despite a wide range of topics, complexities, entanglements and narratives covered in the extracted text, we also inductively found what we subsequently coded and further analysed as ‘doing data discrepancies’. We typified particular dimensions of such discrepancies assigned to certain text passages. Finally, we generated two visualised heuristics that facilitated further refinement of our findings, which we turn to in the next section.

Doing data-based school monitoring in state education agencies: Insights from Massachusetts and Hamburg

While it must be recognised that minor variation exists, the technical infrastructure/process of state school monitoring in Hamburg and Massachusetts can be broadly summarised as follows (Figure 1). Firstly, numerous data points are digitally collected from school and/or student information systems (in the US, with districts acting as data mediators) within varying time frames (from annually to daily). Submitted data is then validated, using a combination of automated and human checking processes, before being centrally stored (either in a data warehouse or in an oracle database). From there, different departmental units make use of data for modelling, analysis and/or data visualisation aligned to different data tools, while also working with external/internal research experts. Finally, the laboriously edited data is widely

reported, both publicly and within different portals used by schools, parents or (in the US) districts.

In general, most of our interviewees were well aware of the complexities behind this technical infrastructure, which might look straightforward on paper, but in practice includes various interdependencies and requires data to flow back and forth multiple times. In fact, most interviewees contrasted their work around data with linear procedures or loop circle models (as the technical infrastructure would suggest), instead describing it as highly experimental, involving significant elements of ‘messing around’ or, as one interviewee phrased it, ‘cooking’ with multiple ingredients (data, algorithms or models) to find working solutions within a highly diverse entanglement of often very different logics, stakeholders or problems.⁵

In line with this argument, interviewees reported that it has become increasingly difficult to organise and work internally with growing amounts of data, which also means an increasing dependence on particular programs, algorithms or indices – with consequent effects due to their selectivity. As one DESE actor in Massachusetts reported:

[I]t’s the program that specifies what you’re doing to the data. It says filter these things out, count that and don’t count this and add these things, but don’t add those things. It’s literally the program, the query that pulls across the data and so on.⁶

Similarly, a German IfBQ actor stated:

An index value always expresses a particular background question, specific method-related consideration.

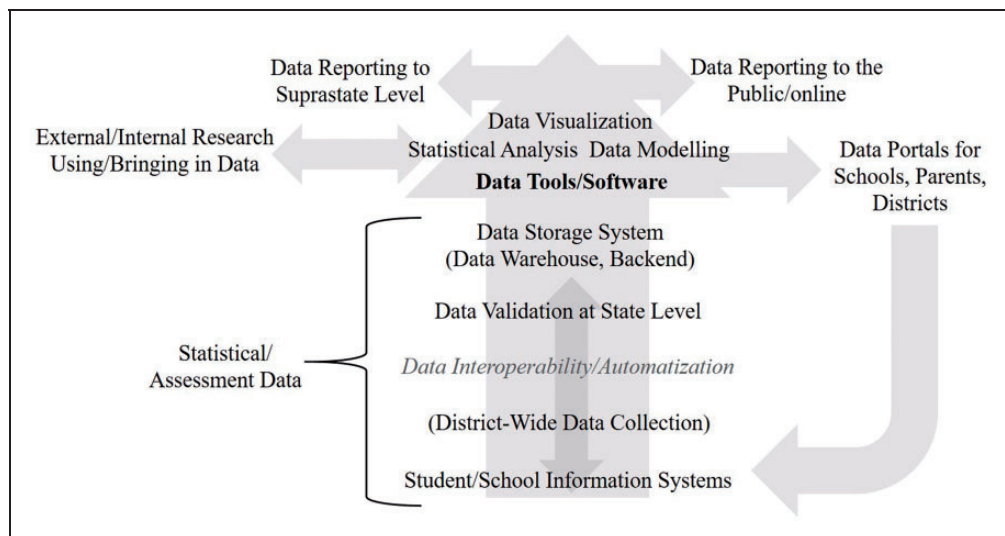


Figure 1. The technical infrastructure of state school monitoring.

And in fact this partly determines how to look at this data.⁷

At the same time, in both countries data practices share a new kind of *data economy*, including attempts to reduce ‘unnecessary’ data, to eliminate data duplication or alternative data expressing the same thing or to automate and accelerate data collection using interoperability standards, centralised platforms and data business rules.

Framed by these more general narratives of both data expansion and data reduction, we identified different discrepancies about which our state agency actors raised concerns when describing their data work, both in a more narrow or technical sense (writing algorithms, building models for calculation, linking data) and also in a wider sense (embedding such technical practices into the wider contexts of school monitoring and leadership) (Figure 2). We discuss the most frequently reported discrepancies, which all share political implications, in greater detail within the following sections.

Data simplification versus data accuracy

A central goal of state education agencies is to nudge schools, teachers, parents or the wider public in the direction of using (their) data more frequently and to improve data-based communication. However, at the same time, our interviewees were well aware that many of their addressees lacked the time, expertise or motivation to dedicate much time to understanding, using and interpreting the (rising amount and complexity

of) data in the ‘right’ way. As one DESE actor in Massachusetts expressed:

What we’re trying to do right now is expand our outreach because we know that there’s a huge opportunity for parents and kids, other audiences to use this data but they’re not going to have as much experience with data, they are not going to be the ones to download it and put it into Tableau and run analytical reports to figure out which school has the best support program.[...] [W]e are trying to [...] really work [...] on data visualisation and doing more actionable data with less interpretation of the data. So that we do it so that parents, we can reach that audience that is not data experts.

Another interviewee from IfBQ in Hamburg made a similar point:

It makes much more sense to arrange data in a way that takes less effort to use. Instead one can instantly draw on it, show things. This [...] map is designed to be printed in any format ready to use for presentations.

As interviewees in both countries reported, the easier and more ready-to-use the data (e.g. using visualisations that tell a clear story), the better it can be ‘understood’ and the more it will also be used by non-experts. We found various examples of such explicitly simplified and condensed data instruments (designed to further prevent schools from ‘drowning in data’, as one agency actor suggested), including one-(web)page data summaries for each school (e.g. the school *Profiles* or

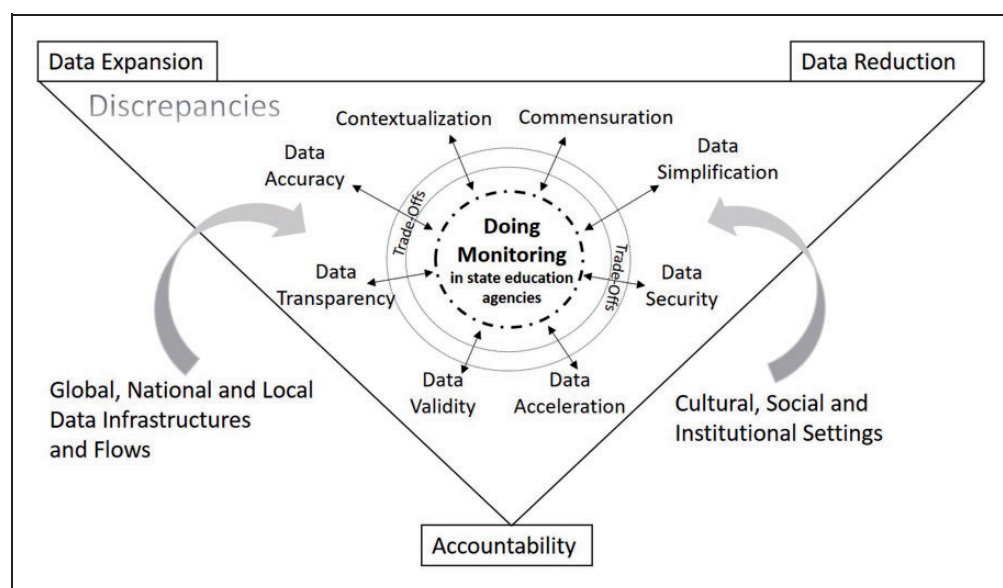


Figure 2. Doing monitoring in state education agencies.

DART (*District Analysis and Review Tools*) instruments in Massachusetts or the *School at a Glance/Schule im Überblick – SchÜb* instrument in Hamburg) as well as the expanding usage of maps, graphs and traffic light systems.

This user-friendly simplification, however, is accompanied by a significant risk of neglecting the multiple possible interpretations of data, intended to be viewed in a context-sensitive manner. In other words, a key issue raised by a number of interviewees was what they described as an unfortunate discrepancy between the demand for simplification and a simultaneous demand for data accuracy, which appeared just as relevant for data communication as for data production and processing.

As an example, the person responsible for the *Early Warning Indicator System*⁸ in Massachusetts said:

Someone was looking at [the data and said] maybe I should use this and start to encourage kids who are at high risk of not going to college or not persisting at college to advise them away from post-secondary. We were like, no! That's exactly the opposite of what this is.

Unsurprisingly, the problem of how to prepare and communicate data with the 'right' amount of complexity increases with every further expansion of potential data audiences that each demand different forms of data preparation, processing and visualisation. For example, an interviewee from Massachusetts said:

I've been a little worried that we have so many data tools, as you've seen [...] and we develop them in so many different ways and we deploy them in so many different ways and I'm worried [...] that we are not clear on what these are all for, who should use which ones, what's the right audience and that kind of thing.

Similarly, an IfBQ actor from Hamburg stated:

[This instrument] is not really user friendly, because it offers something for everybody, which means it actually doesn't offer the right thing for anybody.

Consequently, what many of our respondents aspire to for the future is a further development of customisable, interactive and flexible data instruments, which offer various, user-related options for adaptation, thus communicating the complexities of data without losing attractiveness.

Who to compare?

Closely related to the problem of how (strongly) to contextualise data, a key component of doing monitoring

for school monitoring and leadership lies in commensuration, which is making particular things *as data* comparable to other things *as data*. We found such commensuration practices to be highly relevant for all dimensions of doing monitoring and for the enactment of data infrastructures in state education agencies, while the questions of who is made comparable with whom and which metrics inform the fabrication of comparisons also imply significant challenges.

One challenge comes along with what our respondents reported as the increasing adaptation of the so-called 'fair comparisons'. Different from comparing, for example, a school's performance to the performance of neighbouring schools or a student's performance to peers in his/her class, fair comparisons instead allow us to relate a particular performance result to the schools/students across the state that are the most statistically similar, thus promising a better (fairer) and context-related understanding of data. Such de-territorialised forms of comparison, which have been made possible by data centralisation, interoperability and standardisation, are not limited to measuring and comparing performance data, but instead have increasingly become part of all kinds of data tools used by state education agencies. However, while this growing reference to statistical 'context' has introduced a new ('fairer') dimension of data contextuality, it has also further complicated the question of how much (territorial or statistical) context is needed to properly understand and use data (see last section). In other words, interviewees suggested that it has become increasingly difficult to determine how many and which comparison options should be 'offered' to data users or directly built into data tools.

In Hamburg, we found such challenges reflected in the fabrication of a social index (*Hamburger Sozialindex*) to classify the socio-economic status of schools, which was to be used not only to determine state-provided resources for that school but also to statistically calculate peers for evaluating test performance (then named 'fair comparison'). Interviewees working with this instrument reported that using such an adjusted feedback method can foster data acceptance and can help schools to properly evaluate their own performance. At the same time, however, IfBQ actors also highlighted that school index classifications were developed in 2012 and six years later were outdated, thus causing some schools to feel inadequately represented in data that still informs meaningful decision-making. Responding to this feedback, Hamburg state education agencies for a while (but not anymore) offered an alternative on-demand evaluation to adjust the index data.

In Massachusetts, an instrument reflecting the discrepancy between territorial versus statistical

contextualisation is the so-called Resource Allocation and District Action Reports (RADAR) instrument. RADAR collects and models various district-level data with a focus on improving finances and spending. An important part of RADAR is the option for districts to make sense of their data by comparing themselves to up to 10 other districts, including those which are territorially far away but similar, e.g. in demographics or student performance. However, as one DESE actor reported, getting districts to use different (fairer) forms of data contextualisation has proved challenging:

[...] [M]ost districts will pick at least a couple of comparison districts from those right around them. Then we had thought that our DART list [DART = *District Analysis and Review Tools*, a different data tool on school and district performance] we'll give you our 10 based on demographics, we thought this was great, because it let districts know [...] [there are schools] that they might not realise that were across the state that they had never heard about but are so similar to them. They hated that. They did not consider districts far away to be good comparisons.

Hence, what DESE now does instead is offer different options for either territorial or statistical comparison, again using customisation and flexibilisation to let data users decide who they would like to compare themselves to (see www.doe.mass.edu/research/radar).

Both examples presented highlight the complexity underlying increasing options for commensuration using either territorial or statistical relation-making, or indeed trying to find a balance between the two.

Speeding up data production while improving data validity

Another key problem of doing monitoring within state education agencies is improving data validity while handling (public or political) expectations to produce data more rapidly, frequently or, at best, in real-time. The majority of our interviewees expressed concern about this tension and the challenges of developing solutions to deliver data more quickly or, at least, 'fast enough' (e.g. for publishing educational monitoring reports by the time they are needed to inform decision-making), while still ensuring data quality and validity.

In both cases, state education agencies are dealing with this issue by setting a specific (yet not overly extensive) timeline for data validation practices, including a deadline which defines the moment at which data becomes 'frozen'. In other words, at that point in time (also described as the 'single point of truth') data in the system is perceived to be correct and is further processed into reporting or additional data modelling,

while further data changes are permitted. As one DESE actor described it:

[...] [O]nce they certify it, once every district says okay, I'm certifying this data and we compile it, I think of it as like the big steel door shuts, that's it. [...] You can't go back [...] because once you have that data and you start reporting it out, it's reported out in so many places. So we don't have control of all that anymore.

A similar view was expressed by a BSB actor in Hamburg:

Well, one has to accept that it's actually the right thing not to correct [it] anymore because you already reported the data to the KMK [Standing Conference of the Ministers of Education of the German States] after all, the senator launched the numbers at the press conference. [...] This data shouldn't be changed because it would open up an already published stage of affairs.

Given the importance of this point of literally no return (see exceptions below), state education agencies in both countries are making massive investments in improving data validation within the provided deadline. Such validation includes (both automated and human) data checking practices within the agency,⁹ data business ruling, but also (technical) barriers to prevent schools from reporting incorrect data using error reporting, which, however, has caused a different problem, as one DESE actor described:

What's the bigger problem is [the schools] [...] fix [their data] [...] in order to get the submission through, but they never go back to fix it in their system so that the next time they don't get those errors, is probably the bigger problem.

In other words, state education agencies need to create more interlaying instruments for downstream data validation processes that are also able to remove errors step by step in the source systems. Still, however, strict timelines are reported to complicate this process, as one DESE actor responsible for the collection and processing of assessment data¹⁰ reflected:

So we fix these [...] [data errors] from June while we're scoring the essays over the summer. So it takes four weeks or five weeks to score 2 million essays, 3 million essays we are scoring and know the response from that. While that's going on [...] [the schools are] fixing the data. [...] In September we have official results, and it's almost always perfect.

Despite this growing investment in data validation, data 'mistakes' have not vanished completely. Interestingly,

referring to such cases, several respondents in fact questioned the single point of truth and suggested that ‘wrong’ data should also be corrected after this point, at least if the data is important. For example, one actor from BSB drew the following distinction:

[...] [U]nfortunately one has to say, you only have a certain period of validation. So the data is also looked at in detail afterwards. If something slips through, which regrettably occasionally happens and is very important, the data warehouse and the single point of truth will be changed in retrospect.

Similarly, the assessment expert at DESE said:

[...] [Y]ou live with the one or two errors for school reporting but at the student level we don’t live with it. We still fix it, even now especially for high school because you can’t graduate unless you pass the test. [...] We will go back. Two years later someone says: I know I passed that test, how come I can’t graduate? Okay. So we’ll go down and we’ll do handwriting analysis [...]. For graduation and sometimes for scholarships that we also give because of the tests, we have one person who is always working on forensic examinations.

Unsurprisingly, data that has already been reported (publicly) is significantly less likely to be changed, which in some cases means ‘living with errors’ in order to preserve data coherence.

Increasing both data transparency and data security

A fourth discrepancy which was frequently reported by our interviewees in both countries was the problem of simultaneously increasing demands for data transparency and data security.

Particularly in the US, where a much stronger value has traditionally been placed on publicly available data, respondents strongly supported publishing as much data as possible:

I’m a believer in publishing as much data as we possibly can. I think that the education community, or even just the public at large, has become much more fluent in being able to look at data and understand data. I [...] believe in that you can effect change by just purely publishing data and letting people see it. Because then I think people will start asking questions about it and that’s a form of accountability. We publish an awful lot of data and it’s sort of a philosophy that says, “Let’s put as much out there as we can and let people harvest or digest whatever the appropriate level that’s right for them.” That’s sort of the general idea.

Still, however, actors at DESE were also well aware that some data is perceived to be too sensitive for publication, either for student data security reasons or in order to prevent an unintended overreaction, as one interviewee stated:

There are plenty of people who go overboard and they overreact to the data that they don’t really understand sometimes and they make decisions, even though it’s good data you can make a bad decision with good data.

As two other DESE actors outlined, such overreactions are particularly likely in high stakes contexts, such as student performance results and their effect on schools’ autonomy,¹¹ parental school choice or real estate markets:

[...] [T]he public looks at these and you may think that a 1% difference from the third grade this year and reading results so next year it goes down 1% you would think, so what? That’s statistics. I live in a town where that 1%, people will get concerned about that. I would worry about there’s tons of ways to use this data inappropriately, either tracking them or shaming students and worry about some of that. So I want to make sure that folks who are using it are using it in a way that helps them.

Consequently, also US interviewees reported that they were extremely careful in deciding which data is made available to whom (often using data security roles for selective database access), particularly in the case of parents:

We don’t have a parent portal, that’s always been talked about, there are lots of really tactical implications around how to ensure that the right people have access to it.

Given that Germany has been much more sceptical about publishing data (which is still more strongly regulated than in the US), the relation between published data (e.g. the aforementioned social index) and for example parental school choice is much less prominent, as one IfBQ actor explained:

One could suspect that [relation]. We never really examined that. But these are topics and things [school choice decisions] that are mostly talked over privately, unrelated to the social index. Not even all parents know about that instrument. [...] Parents have a particular idea of school anyways [...] one could assume that some aspects of parental school choice are influenced by that. But that is something you would have to take a closer look at.

Nevertheless, our respondents from Hamburg also expressed a growing pressure to make data available and cited schools worrying about the potential future usage of their data for accountability purposes:

As soon as schools notice something is done with their data, you should carefully consider whether you do it or not. [...] Schools are quite sensitive in this regard.

Finally, in both countries, given the growing prominence of data-based school governance, data security has become an increasingly political issue, resulting in new data protection laws and regulations. As one IfBQ actor said:

Data protection has become much more important over the past few years, compared to 10-15 years ago, because there are many more possibilities and there is much more data available than a few years ago. Hence, it's important to describe methods and processes very clearly, to define the conditions for linking data.

In both countries, a common way of dealing with these ambivalent requirements has been to introduce complex systems of data pseudonymisation, which allows a great deal of (linked) data to be processed and published without identification.

Proactively considering the ambivalent effects of accountability

As mentioned in the case selection part of this paper, the US and Germany stand in stark contrast in terms of their attitudes towards using and relying on (quantified) data – particularly for accountability purposes – with Germany being much more sceptical towards published data, standardised testing and the use of high-stakes rankings. Having said this, while using data for accountability was still much less of an issue in Hamburg (yet this seems to be gradually changing, see below); in Massachusetts it was frequently mentioned as a key feature of doing data, which simultaneously, however, appears to intensify the aforementioned discrepancies. DESE actors were thus well aware that using data to enforce accountability and build accountability models is strongly influenced by the norms and values used as underlying benchmarks:

[...] [W]e've done some pretty deep philosophical discussions when we are debating what indicators to include, how much improvement we should expect to see and it's an interesting balance of this technical side and then this normative side. Because in the end, you've got to say, did this school make it or not?

Modelling accountability (data) consequently poses additional challenges, as, on the one hand, the data and models are still intended to be used by schools for their own improvement. On the other hand, the stakes attached to accountability data enhance new dynamics related to the model, such as 'blind' performativity or gaming. Several DESE actors expressed concerns about schools manipulating data in response to accountability measures, with such a risk strongly affecting how they do data in this field:

From the accountability perspective, it could be tricky to put [particular] [...] kind[s] of measures in accountability. Like, for example, we're piloting a school climate survey. We could down the road consider putting that in, but you create these centres for teachers or principals or whoever to tell the kids, make sure you fill all those out, the top possible score in a way that's harder to do with an assessment, like it's harder to manipulate assessment graduation rates. So that's where I think the conversation gets more challenging, is around using accountability.

As mentioned above, despite the officially remaining scepticism against high-stakes accountability, even in Germany we found that the rising amount of available data and the use of that data for resource distribution, state-school target agreements or school support (via 'close consultation' and 'continuous data feedback', as one actor described it) has in fact significantly increased the use of data for accountability purposes (which is not simply low-stakes anymore). Two statements nicely illustrate this visible discrepancy. One IfBQ actor said:

Political decisions won't be linked to that data [...]. Not like in the US where schools can be closed [based on accountability scores]. [...] We don't think about this. And I think that's good.

Later in the interview, the same actor said:

There are a lot of schools that, from our view, are doing a good job on [...] [using data]. They look at the results, discuss them [...], build on them for school development. Schools have also got their target agreements with their school supervision agency, where such topics are discussed as soon as they notice that they perform lower than comparable schools. They say they want to change something, want to improve some student groups who perform badly. Many schools do a great job on that.

Finally, in Hamburg, interviewees also expressed concern about schools gaming data collections that are

linked, in particular, to resource distribution, such as the aforementioned social index. For example, interviewees at the IfBQ reported that some schools were worrying that other schools could manipulate their social index by reporting more students with low social-economic status than were actually attending in order to receive more funding.

Concluding remarks

The aim of this paper was to provide empirical insights on how state education agencies in Germany and the US enact the rising datafication of schooling, looking, in particular, on data infrastructures, flows and practices for school monitoring. Our findings have thus illustrated some selected features of this ‘doing’ monitoring as reported by actors from three state education agencies in two different national contexts. Even though our findings generally reveal what Selwyn (2013: 198) has described as the ‘messy’ realities of technology and education, we still identified different types of ‘doing data discrepancies’ that present somewhat typical challenges and ambivalences described by our interviewees in both countries in surprisingly similar ways. This is despite the fact that our respondents from Hamburg continue to articulate strong criticism of (public) rankings and high-stakes testing. Nonetheless, we found contextuality to be highly reflected in the doing of data, for example as evident in the example of making data publicly available. It is important to mention once more that both selected cases – *within* their country context – represent advanced states with a comparatively (in relation to their state peers) long history of datafication and an extensive use of data for monitoring and school leadership. However, since we also found many of our findings reflected in other states (both in Germany and the US), we assume our conclusions to be of significant cross-context relevance.

At the same time, the discrepancies reported by our interviewees show how the social and technical are not only deeply interwoven with data-based school monitoring, but also – as emphasised in the existing critical data studies literature – how data practices always have political implications, particularly when applied to systems of (high *or* low stakes) accountability. As Kitchin and Lauriault (2014: 4–5) state, data infrastructures are always ‘[...] expressions of knowledge/power, shaping what questions can be asked, how they are asked, how they are answered, how the answers are deployed, and who can ask them’ (see also Ruppert et al., 2017). In other words, monitoring infrastructures create what West (2017: 1) describes as limited ‘[...] second-hand representations of important objects of analysis’ that administrators use to speak on behalf of the school, the teacher or the student. As our findings illustrate,

such representations not only create new categories (e.g. students ‘at risk’ or schools ‘in need’) but make things (in)visible or predictable in particular ways, ultimately having the potential to change ‘[...] the essential qualities of what is being studied’ (West, 2017: 10–11). While new information thus continuously produces the need for more and better information (Thompson and Sellar, 2018), it simultaneously increases the amount of data management, including data about data production (evident for example in the expansion and evaluation of data validation and data business rules) (Piattoeva, 2016), ultimately shifting more and more attention towards ‘valid’, ‘fast’ or ‘usable’ representation-making.

Nonetheless, a key result of our study is that datafication, at least in our selected state education agencies, does not appear to produce single centres of calculation and data power, but is instead mediated through multiple infrastructures and practices that together perform calculation, commensuration and data work. Against this backdrop, we fully agree with Gray et al. (2018: 1) that instead of (only) calling for data literacy in the sense of competencies in reading and working with datasets, there is a pressing need for so-called *data infrastructure literacy*, which is ‘[...] the ability to account for, intervene around and participate in the wider socio-technical infrastructures through which data is created, stored and analysed’. This, however, requires close empirical observation of data infrastructures at work, not only in the field of education but also with regard to wider issues of governance, data-driven policy-making and the organisation of the state. In this regard, our study, and particularly our attempt, to both visualise and (at least partly) typify the data infrastructure of doing monitoring in state education agencies, while specific in its contribution, might also be applicable to other policy fields.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The presented research was funded by the *Deutsche Forschungsgemeinschaft* [DFG, German Research Foundation] – project number HA 7367/2-1.

Notes

1. We use the term ‘state’ here to describe subnational units of educational authority, as commonly used in the US. It seems important to note that such subnational authorities are usually named *Länder* in Germany. Yet, we use the

- US term here for the purpose of alignment between the cases.
- For a closer examination of the policy context around expanding datafication in both countries see Hartong (2018b).
 - Even in the US, where districts have traditionally been charged with monitoring schools through data, state education agencies have increasingly become the main source for educational data because they not only collect data from every student and teacher in every school and every district of the state but can also make comparisons using these data.
 - The interviews we draw on for this contribution are part of a larger project, funded by the *German Research Foundation* (DFG, project number HA 7367/2-1), which seeks to improve the understanding of digital-era governance and the role of data management in education within the policy contexts of Germany and the US. The project includes analyses of (1) policy material, such as monitoring regulations, resolutions or digitalisation/datafication programmes, (2) the actors and institutions involved in performance data management at national and state level, using policy network analysis, (3) the performance data infrastructures and their modes of operation in selected state education agencies in each country, as well as (4) interviews with national and state-level policy actors, technicians, administrators and data system providers.
 - At the same time, several interviews clearly pointed out that there is no and will never be 'perfect data' mirroring 'the truth'. As one interviewee stated: 'Any dataset that you think is perfect, you just haven't looked at it close enough'.
 - To aid legibility and understanding, interview quotes were slightly altered both linguistically and grammatically.
 - German interview transcripts were translated into English by the authors.
 - For an overview of these kinds of data tools, see Neild et al. (2007).
 - Interviewees reported that this checking process is usually directed at finding non-sense data (e.g. a seven-year-old student attending 12th grade), data doubles, data inconsistencies or missing data.
 - The assessments are increasingly carried out online, so much more data is collected automatically. However, free text parts are still coded manually.
 - As an example, schools performing poorly in accountability measures can be taken over at least temporarily by DESE, see also below.
- Bellmann J (2015) Symptome der gleichzeitigen Politisierung und Entpolitisierung der Erziehungswissenschaft im Kontext datengetriebener Steuerung. *Erziehungswissenschaft* 26(50): 45–54.
- Bloem S (2016) *Die PISA-Strategie der OECD: zur Bildungspolitik eines globalen Akteurs*. Weinheim: Beltz Juventa.
- Diesner J (2015) Small decisions with big impact on data analytics. *Big Data & Society* 2(2): 1–6.
- Espeland WN and Stevens ML (1998) Commensuration as a social process. *Annual Review of Sociology* 24(1): 313–343.
- González-Sancho C and Vincent-Lancrin S (2016) Transforming education by using a new generation of information systems. *Policy Futures in Education* 14(6): 741–758.
- Gorur R (2014) Towards a sociology of measurement in education policy. *European Educational Research Journal* 13(1): 58–72.
- Gray J, Gerlitz C and Bounegru L (2018) Data infrastructure literacy. *Big Data and Society* 5(2): 1–13.
- Grek S and Ozga J (2010) Governing education through data: Scotland, England and the European education policy space. *British Educational Research Journal* 36(6): 937–952.
- Hartong S (2016) Between Assessments, Digital Technologies, and Big Data: The Growing Influence of 'Hidden' Data Mediators in Education. *European Educational Research Journal* 15(5): 523–536.
- Hartong S (2018a) Towards a topological re-assemblage of education policy? Observing the implementation of performance data infrastructures and 'centers of calculation' in Germany. *Globalisation, Societies and Education* 16(1): 134–150.
- Hartong S (2018b) The transformation of state monitoring systems in Germany and the US: relating the datafication and digitalization of education to the Global Education Industry. In: Parreira do Amaral M, Steiner-Khamsi M and Thompson C (eds.) *Researching the Global Education Industry: Commodification, the Market and Business Involvement*. Palgrave Macmillan, pp. 157–180.
- Hartong S (2019) Politikmobilität und 'datenbasierte' Educational Governance: (Weiter-) Entwicklung einer topologischen Perspektivierung. *Bildung und Erziehung* 72(1): 6–23.
- Hopmann S (2003) On the evaluation of curriculum reforms. *Curriculum Studies* 35(4): 459–478.
- Iliadis A and Russo F (2016) Critical data studies: An introduction. *Big Data & Society* 3(2): 1–7.
- Kitchin R (2014) *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. London: SAGE.
- Kitchin R and Lauriault T (2014) Towards critical data studies: Charting and unpacking data assemblages and their work. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2474112 (accessed 16 May 2019).
- Landri P (2018) *Digital Governance of Education. Technology, Standards and Europeanization of Education*. London: Bloomsbury Academic.
- Lewis S (2017) Governing schooling through 'what works': The OECD's PISA for Schools. *Journal of Education Policy* 32(3): 281–302.

ORCID iD

Sigrid Hartong  <https://orcid.org/0000-0002-8618-4837>

References

- Anagnostopoulos D, Jacobsen R and Rutledge A (eds) (2013) *The Infrastructure of Accountability. Data Use and the Transformation of American Education*. Cambridge: Harvard Education Press.

- Lingard B, Martino W, Rezai-Rashti G, et al. (2015) *Globalizing Educational Accountabilities*. London: Routledge.
- Neild RC, Balfanz R and Herzog L (2007) An Early Warning System. *Educational Leadership* 65(2): 28–33.
- Piattoeva N (2016) The imperative to protect data and the rise of surveillance cameras in administering national testing in Russia. *European Educational Research Journal* 15(1): 82–98.
- Piattoeva N, Centeno VG, Suominen O, et al. (2018) Governance by data circulation? The production, availability, and use of national large-scale assessment data. In: Kauko J, Rinne R, Takala T (eds) *Politics of Quality in Education. A Comparative Study of Brazil, China, and Russia*. London: Routledge, pp. 115–136.
- Rivera MD (2018) Paying for financial expertise: Privatization policies and shifting state responsibilities in the school facilities industry. *Journal of Education Policy* 33(5): 704–737.
- Ruppert E, Isin E and Bigo D (2017) Data politics. *Big Data & Society* 4(2): 1–7.
- Sacks P (1999) *Standardized Minds: The High Price of America's Testing Culture and What We Can Do To Change It*. New York: Perseus Books.
- Schildkamp K and Teddlie C (2008) School performance feedback systems in the USA and in The Netherlands: A comparison. *Educational Research and Evaluation* 14(3): 255–282.
- Sellar S (2015) Data infrastructure: A review of expanding accountability systems and large-scale assessments in education. Discourse. *Studies in the Cultural Politics of Education* 36(5): 765–777.
- Sellar S (2017) Making network markets in education: The development of data infrastructure in Australian schooling. *Globalisation, Societies and Education* 15(3): 341–351.
- Selwyn N (2013) Rethinking education in the digital age. In: Orton-Johnson K and Prior N (eds) *Digital Sociology: Critical Perspectives*. Berlin: Springer, pp. 197–212.
- Selwyn N (2014) Data entry: Towards the critical study of digital data and education. *Learning, Media and Technology* 40(1): 1–18.
- Smith GJD (2018) Data doxa: The affective consequences of data practices. *Big Data & Society* 5(1): 1–15.
- Sobe NW and Kowalczyk J (2018) Context, entanglement and assemblage as matters of concern in comparative education research. In: McLeod J, Sobe NW and Seddon T (eds) *World Yearbook of Education 2018: Uneven Space-Times of Education: Historical Sociologies of Concepts, Methods and Practices*. London: Routledge, pp. 197–204.
- Thompson G and Sellar S (2018) Datafication, testing events and the outside of thought. *Learning, Media and Technology* 43(2): 139–151.
- Villani M (2018) The production cycle of Pisa data in Brazil. The history of data beyond the numbers. *Sisyphus* 6(3): 30–52.
- West J (2017) Data, democracy and school accountability: Controversy over school evaluation in the case of DeVasco High School. *Big Data & Society* 4(1): 1–16.
- Williamson B (2017) *Big Data in Education: The Digital Future of Learning, Policy and Practice*. Los Angeles: SAGE.
- Williamson B, Bergviken Rensfeldt A, Player-Koro C, et al. (2018) Education recoded: Policy mobilities in the international 'learning to code' agenda. *Journal of Education Policy*. Advance online publication. DOI: 10.1080/02680939.2018.1476735.